

## Case Report ■

# Developing Regulatory-compliant Electronic Case Report Forms for Clinical Trials: Experience with The Demand Trial

BOGDAN ENE-IORDACHE, ENG D, SERGIO CARMINATI, IT, LUCA ANTIGA, PH D, NADIA RUBIS, RN, PIERO RUGGENENTI, MD, GIUSEPPE REMUZZI, MD, ANDREA REMUZZI, ENG D

**Abstract** The use of electronic case report forms (CRF) to gather data in randomized clinical trials has grown to progressively replace paper-based forms. Computerized form designs must ensure the same data quality expected of paper CRF, by following Good Clinical Practice rules. Electronic data capture (EDC) tools must also comply with applicable statutory and regulatory requirements. Here the authors focus on the development of computerized systems for clinical trials implementing FDA and EU recommendations and regulations, and describe a laptop-based electronic CRF used in a randomized, multicenter clinical trial.

■ J Am Med Inform Assoc. 2009;16:404–408. DOI 10.1197/jamia.M2787.

## Introduction

This report describes the research group's approach to the problem of developing regulatory-compliant computerized CRF systems for use in randomized trials. In medical research, controlled clinical trials establish the efficacy (or lack thereof) of medications and clinical interventions; for many research questions, they provide the most reliable form of scientific clinical evidence.<sup>1</sup> Clinical trial data collection and management consumes substantial resources in time and effort, especially when trials rely heavily on paper documents. Numerous reports indicate that the efficiency can be improved by replacing paper CRF with electronic ones (e-CRF).<sup>2,3,4</sup> With the expansion of Internet-based applications in many fields, the demand for e-CRF is growing.<sup>5</sup> The purpose of this report is to review regulatory requirements regarding e-CRF development and implementation, and to describe an example system that complies with the requirements.

## Case Description

Since EDC has become a valid alternative to paper-based trials, developers face the challenge of creating tools compliant with FDA and European regulations. The rules of *Good Clinical Practice* (GCP), which dictate the principles of

data integrity in paper-based data collection, must apply equally to EDC. The development and use of computerized systems for clinical trials are specifically regulated. In the United States the *Electronic Records, Electronic Signature* regulation, known as 21 CFR Part 11,<sup>6</sup> was published in 1997 to address issues for quality, security and integrity of data that the FDA will accept as equivalent to paper records. Subsequently, the FDA published a Guidance for computerized systems used in clinical trials in 1999,<sup>7</sup> that was reorganized and republished in 2007.<sup>8</sup> In Europe the requirements for e-CRF are included in the GCP guidelines of the European Medicines Agency (EMA), published in 2002.<sup>9</sup>

**Regulatory Requirements.** The 21 CFR Part 11<sup>6</sup> rules state that computerized systems should meet all regulatory requirements with the same quality as paper-based data collections and must use electronic signatures as the legally binding equivalent of individual handwritten signatures. The electronic system designs must satisfy all requirements of the study protocol (e.g., blinded study, cross-over, etc). Detailed requirements for e-CRF, as provided by the FDA and EU<sup>8,9</sup> Guidances are presented in Table 1. To achieve regulatory compliance, systems and research projects employ procedural and technical controls. Procedural controls involve administrative actions, dealing with trial organization and documentation. Technical controls comprise measures that ensure the quality, accuracy, and integrity of data stored in the electronic systems. These can be grouped by the type of feature covered, into several categories (see Table 1). User **authorization** controls are security measures to identify the person who submits the data, to prevent unauthorized access to the system. **Audit trail** controls are measures to ensure that the system keeps a record about sources from which data originates, who made changes, when, and what information was changed. **Attributability** controls are measures to ensure that data will be retrievable in such a way that all information regarding each subject in a study is attributable to that subject. **Data validation** controls are checks performed by the computerized systems to ensure the validity and quality of clinical information. **System**

Affiliations of the authors: Clinical Research Center for Rare Diseases, Aldo e Cele Daccò, Mario Negri Institute for Pharmacological Research (BE-I, SC, LA, NR, PR, GR, AR), Ranica (BG), Italy; Unit of Nephrology, Azienda Ospedaliera Ospedali Riuniti (PR, GR), Bergamo (BG), Italy; Department of Industrial Engineering, Università di Bergamo (AR), Dalmine (BG), Italy.

The authors are grateful to all the investigators and monitors who participated in the DEMAND trial. The authors greatly appreciate the help of J.D. Baggott in preparing the manuscript. The authors thank Chiesi Farmaceutici Spa, Parma, Italy for their support for the DEMAND trial.

Correspondence: Bogdan Ene-Iordache, Eng D, Laboratory of Biomedical Technologies, Clinical Research Center for Rare Diseases Aldo e Cele Daccò, Via G.B. Camozzi 3, 24020 Ranica (BG), Italy; e-mail: <bodgan@marionegri.it>.

Received for review: 03/12/08; accepted for publication: 01/31/09

Table 1 ■ Regulatory Requirements for Electronic Trial Data Systems

Action (Feature)		FDA Guidance <sup>8</sup>		EMA Guidance <sup>9</sup>
Procedural	A	Study Protocols. Each specific study protocol should identify each step at which a computerized system will be used to create, modify, maintain, archive, retrieve, or transmit source data.	6.4.9	The clinical trial protocol should contain the identification of any data to be recorded directly on the CRFs (i.e., no prior written or electronic record of data), and to be considered source data.
Procedural	B	There should be sops and controls in place when using computerized systems to create, modify, maintain, or transmit electronic records, including when collecting source data at clinical trial sites.	5.5.3.b	Maintain sops for electronic systems.
Procedural	C	When original observations are entered directly into a computerized system, the electronic record is the source document. under 21 CFR 312.62, 511.1(b) <sup>7</sup> (ii) and 812.140, the clinical investigator must retain records required to be maintained under part 312, <sup>8</sup> 511.1(b), and part 812, for a period specified in these regulations.	4.9.4	The investigator/institution should maintain the trial documents as required by the applicable regulatory requirement(s). The investigator/institution should take measures to prevent accidental or premature destruction of these documents.
Technical (Authorization)	D1	Access must be limited to authorized operators (21 CFR 11.10)(d) that have an individual account. The user should always log out at the completion of data entry session or when leaving the workstation. Alternatively, an automatic log off may be appropriate.	5.5.3.d	Maintain a security system that prevents unauthorized access to the data.
Technical (Audit Trail)	D2	Keep track of all changes made to information in the electronic records that document activities related to the conduct of the trial (audit trails). Audit trails or other security methods used to capture electronic record activities should describe when, by whom, and the reason changes were made to the electronic record.	5.5.3.c	Ensure that the systems are designed to permit data changes in such a way that the data changes are documented and that there is no deletion of entered data (i.e., maintain an audit trail, data trail, edit trail).
Technical (Audit Trail)	D3	Ensure that the system's date and time are correct. The ability to change the date or time should be limited to authorized personnel.		
Technical (Authorization)	E	External safeguards to ensure that access to the computerized system and to the data are restricted to authorized personnel. Prevent the altering, browsing, querying, or reporting of data via external software applications that do not enter through the protective system software.	5.5.3.e	Maintain a list of the individuals who are authorized to make data changes.
Technical (Data Validation)	F1	Incorporate features into the computerized system to encourage consistent use of clinical terminology and to alert the user to data that are out of acceptable range.		
Technical (Attributability)	F2	The computerized system should be designed in such a way that retrieved data regarding each individual subject in a study is attributable to that subject.		
Procedural	F3	Documentation should identify what software and hardware will be used to create, modify, maintain, archive, retrieve, or transmit clinical data.	5.5.3.b	Maintain sops for electronic systems.
Technical (System Integrity)	F4	Sufficient backup and recovery procedures should be designed to protect against data loss.	5.5.3.f	Maintain adequate backup of the data.
Technical (System Integrity)	F5	Integrity of the data and the integrity of the protocols should be maintained when making changes to the computerized system, such as software upgrades, including security and performance patches, equipment, or component replacement.		
Procedural	G	Training should be provided to individuals in the specific operations with regard to computerized systems that they are to perform.		

See<sup>8</sup> and<sup>9</sup> for details on regulatory recommendations.

**integrity** measures are all the procedures to guarantee the integrity of the system and protect against data loss.

Here the authors focus mainly on the technical features that entities must address when they develop, maintain and use electronic record systems for clinical trials, providing the example of a laptop-based e-CRF that the authors used for data management in the DEMAND (delapril and manidipine in Diabetes) trial.

## Methods

While it is challenging to implement the technical features presented above in a single application, several commercial and open-source clinical research data management systems have done so.<sup>14–19</sup> The authors have developed a regulatory-compliant system that has a somewhat typical architecture, with specific components as described below.

An efficient approach employs several components, each with specific tasks, working jointly as a whole e-CRF. The typical architecture of a client-server system, presented in Fig 1A consists of an operating system (OS), a relational database management system (DBMS) for clinical data storage, and a graphic user interface (GUI), namely the e-CRF application. As shown in Fig 1B they may reside on different hosts in a network, but in a scenario for stand-alone laptops they must reside on the same computer. The components presented in Fig 1 are universal, i.e., the OS can be any one, the DBMS can be any of the relational database servers, and the GUI can be implemented in many ways.

The authors addressed user **authorization** carefully in the DEMAND trial e-CRF. All operators signed a statement certifying that their electronic signature, defined by a private

pair of username and password, had the same legal significance as their traditional handwritten signature. To ensure that operators had the right to perform certain actions, a privilege system was implemented by defining user roles (clinical investigators, clinical monitors, and system administrators) and secured log-in at both the OS and GUI levels. As recommended by point D1<sup>8</sup> (see Table 1), the current user, database accessed, date, and time are always highlighted on the GUI status bar. The application is stopped automatically after 10 minutes of user inactivity.

The authors implemented the **audit trail** of the e-CRF using the enterprise features of the DBMS. The back-end database included tables for clinical data and supporting tables for user management and audit trail. Database triggers were implemented for auditing any inserted, updated or deleted data. Every event is stored in the audit table together with the timestamp, the user ID of the person doing the operation, the old and new values of the changed field.

The **attributability** of information was ensured at the levels of DBMS and GUI. The authors designed the clinical tables according to the study flow chart, setting the screening number as the principal identifier of subjects. Queries combined with stored procedures, implemented also at the GUI level, allow easy browsing of clinical information and very fast retrieval of individual data by visit (see Fig 2). The authors set up the data **validation** at the level of GUI. Specific checks were implemented to guarantee validity of data as per study protocol (existence of date of birth, sex, and informed consent at the time of new patient insert, inclusion and exclusion criteria, randomization), warnings for values outside the normal range, and calculated fields.

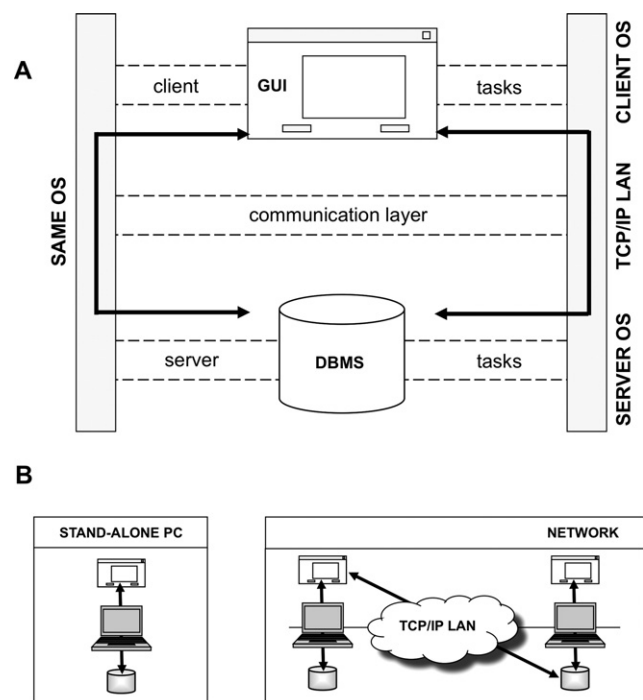
The **system integrity** was maintained by the system administrator. All database instances were scheduled for daily backup and additional backups were done on the server and on securely maintained external devices. Antivirus protection software was installed on all computers and on the network firewall.

It took approximately the equivalent of one developer working six months full time for system development. The e-CRF was designed for academic research only. There is no patent pending and no financial conflict of interest between the work of the authors and the implementation of this computerized system.

## Example

*The DEMAND Trial.* DEMAND was a prospective, randomized, double-blind, parallel-group trial aimed at preventing the onset of nephropathy in hypertensive, type 2 diabetic patients. The trial was coordinated by the Mario Negri Institute (MNI), Bergamo, Italy, in cooperation with seven diabetes outpatient clinics. As these were all located near Bergamo city, the authors decided to adopt an e-CRF based on laptop computers, one for each center. Recorded clinical information included baseline demographics and medical history, and follow-up blood and urine tests, concomitant medication, diseases, and adverse events.

*Study Management.* Investigators were physicians authorized and trained to use the e-CRF at the MNI. Each investigator was provided with a laptop. Monitoring and study drug management was also done by MNI staff autho-



**Figure 1.** (A) All-purpose client-server architecture for e-CRFs: a relational DBMS provides clinical data storage; the client (GUI) can communicate with the DB either locally or through the network. (B) DBMS and GUI can reside on the same OS (stand-alone) or on different OS (networked).

**LABORATORY EXAMINATION / SCREENING #: 3000**

This record is locked!

VISIT 04 DATE (dd/mm/yyyy) 24/05/2002 SCREENING # 3000

Date of sample (dd/mm/yyyy) 24/05/2002 Was the patient fasting or not? 1 1=Yes; 0=No; 2=NA

CREATININE (mg/dl) 1.08	GLUCOSE (mg/dl) 185	LEUKOCYTES ( $\times 10^3/\mu\text{l}$ ) 5.0
UREA (mg/dl) 46	HBA1C (%) 5.37	PLATELETS ( $\times 10^3/\mu\text{l}$ ) 160
SODIUM (mEq/l) 137.9	FRUCTOSAMINE (mmol/l) 2	ERYTHROCYTES ( $\times 10^6/\mu\text{l}$ ) 5.06
POTASSIUM (mEq/l) 4.34	URICACID (mg/dl) 7.0	HEMOGLOBIN (g/dl) 14.8
GOT (AST) (U/l) 19	T. CHOLESTEROL (mg/dl) 216	HEMATOCRIT (%) 44.7
GPT (ALT) (U/l) 22	HDL (mg/dl) 57.0	MCV (fl) 88.4
ALKALINE PHOSPHATASE (U/l) 37	(calculated) LDL (mg/dl) 159.0	
GAMMA GT (U/l) 29	TRIGLYCERIDES (mg/dl) 135	
TOTAL BILIRUBIN (mg/dl) 0.91	TOTAL PROTEINS (g/dl) 6.6	
	ALBUMIN (g/dl) 4.1	

Visit	Sample Date	Cre	Urea	Na	K	Glucose
01	07/03/2002	0.9	42	137.6	4.24	
04	24/05/2002	1.08	46	137.9	4.34	
05	31/05/2002	1.02	43	136.3	4.17	
06	21/06/2002	1.19	39	137.7	4.19	
08	30/08/2002	1.02	41	139.2	4.19	
09	22/11/2002	1.11	39	133.7	3.76	
10	21/02/2003	1	47	136.9	4.41	
11	27/05/2003	0.99	45	135.6	3.94	

Buttons: Add, Update, Delete, Refresh, Close

Status: Current User: mrossi Current DB: VILL\_S 06/03/2008 15:05

**Figure 2.** A representative screenshot of the GUI: master-detail form for laboratory data entry. The user can enter or modify data for the current visit in the top part of the form. In the bottom part there is a grid summarizing all other visits. Navigation between visits is possible using the arrows at the bottom of the form or by clicking on the grid rows.

alized to use the e-CRF on the laptops. The hardware for the DEMAND trial comprised eight laptops, two desktop PCs, and one dedicated server. The system included a Windows® 2000 Server with Windows® XP Professional clients. For clinical data storage the authors used MaxDB, a powerful enterprise database server developed by SAP (SAP AG, Walldorf, Germany). Despite being an enterprise-class DBMS, MaxDB is not restricted to server-based situations, but can be used as a desktop or laptop DBMS as well.<sup>10</sup> This means laptop computers can work either in stand-alone or network mode. To develop the GUI the authors used Visual Basic integrated development environment and ADO (Activex data objects) for database programming. The GUI was designed as a multiple document interface (MDI) model, with a fixed subject ID list on the left and floating child windows as data entry forms.

A representative screenshot of the GUI is presented in Fig 2. The main features of the e-CRF include automatic generation of sequential screening numbers, clinical data capture, fast browsing of stored information, standardized coding of diseases, medications and adverse events.<sup>11,12</sup>

**Trial Database Status.** Patient recruitment for the DEMAND study was completed between September 2002 and November

2005. To export all the clinical data of the DEMAND trial, databases from single laptops were first copied onto the domain server. Then, clinical tables were exported as text files with a python<sup>13</sup> script designed to append data in unified datasets. The characteristics of the trial database are presented in Table 2. Nine hundred nine Caucasian patients with type 2 diabetes were screened, and 492 were enrolled into the study. Of these, 380 (77%) were randomized to the study treatments. Clinical investigators made 6,745 visits,

**Table 2 ■ Status of the DEMAND Trial Database**

Number of patients in database	909
Number of patients enrolled	492
Number of visits	6,745
Mean visits per patient	14
Number of patients randomized	380
Number of tables—clinical data	28
Mean number of fields per table—clinical data	21
Number of clinical data records	53,378
Number of tables—other data	2
Total records—other data	157,028
Total size of the database*	550 Mbytes

\*For 8 MaxDB instances, one for each participating center.



corresponding to an average of 14 per subject. The database consisted of 28 tables for clinical information, for a total of 53,378 records and two supporting tables for user role administration and the audit trail, which had 157,028 records. With this amount of data the database took up 550 megabytes of physical disk space.

## Discussion

Since agency acceptance of data from electronic trials for decision purposes depends on the ability to verify the quality of the data, computerized systems must follow regulatory recommendations. Our in-house developed e-CRF is an example of implementation of the technical requirements recommended by the FDA<sup>8</sup> for computerized systems for clinical trials.

Although it was tailored for the DEMAND trial, the e-CRF in our example could be easily applied in general to randomized clinical trials, or the regulatory compliance principles ported to other existing systems, preserving the system architecture presented in Fig 1. Since part of the collected data (i.e., demographics, medical history, adverse events) are common to many clinical studies, some software modules are re-usable, while for trial specific data (inclusion criteria, blood tests, drug supply, etc) writing new code for DB tables and GUI would be necessary. Based on experience to date, the authors estimate that setting up a new e-CRF will require one person to work full time for 1 month.

The authorization, attributability, and audit trail features at OS or DBMS level resulted in a fast and reliable e-CRF. Generating screening numbers based on DBMS sequences and data validation rules implemented at the GUI level guaranteed high quality of clinical data. The system proved very flexible for clinical data handling, either for DEMAND trial investigators or monitoring staff.

Internet-based trials have gained popularity due to their ability to cover world-wide, multisite projects. Researchers may choose from several commercial solutions from big data warehouse vendors,<sup>14,15,16</sup> or may prefer an open-source Web-based solution, like OpenClinica,<sup>17</sup> DADOS<sup>18</sup> and REDcap.<sup>19</sup> Our e-CRF is not currently available on an open-source basis but those interested in more detail regarding implementation of the system may contact the authors. Even if the implementation platform is different (e.g., laptop-based or Internet-based), it is interesting to see how our e-CRF compares with these other systems. For example, the client-server architecture proposed in Fig 1 appears valid for these systems, even if the server is a web server and uses an Internet browser as GUI. The implementation of the audit trail in OpenClinica<sup>17</sup> is similar to ours, being based mainly on database-level triggers. Our e-CRF has several advantages of a client-server system that can work either in stand-alone or network mode: data are always available, client GUI browse data quickly, and built-in OS features for user and software management can be used. By contrast, our system does not allow for the easy implementation of new trials, cannot be used for remote data entry and requires heavier maintenance than a web-based system. Some inter-

esting features that the system lacks, like data encryption and integration with other medical record systems, might be implemented in future developments.

Along with common features like EDC, data validation, user management, audit trail and data export that are common to all types of e-CRF, our approach demonstrates the harmonization of these components, from the OS to the DBMS and the GUI, to form a single e-CRF that implements the regulatory features required for computerized systems used in clinical investigations. The results further support the use of information technology in this strategic area of drug research and development.

## References ■

1. Kunz R, Oxman AD. The unpredictability paradox: Review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* Oct 31 1998;317(7167):1185–90.
2. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials* 2008;5(1):75–84.
3. Lopez-Carrero C, Arriaza E, Bolanos E, et al. Internet in clinical research based on a pilot experience. *Contemp Clin Trials* Apr 2005;26(2):234–43.
4. Pepine CJ, Handberg-Thurmond E, Marks RG, et al. Rationale and design of the international verapamil SR/Trandolapril Study (INVEST): An internet-based randomized trial in coronary artery disease patients with hypertension. *J Am Coll Cardiol* Nov 1998;32(5):1228–37.
5. Chaudhry B, Wang J, Wu S, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* May 16 2006;144(10):742–52.
6. FDA. 21 CFR Part 11: Electronic records; electronic signatures; final rule. *Fed Regist* 1997;62(54):13429.
7. FDA. Guidance for industry: Computerized systems used in clinical trials, 1999.
8. FDA. Guidance for industry: Computerized systems used in clinical investigations, 2007.
9. EMEA; ICH. Topic E 6 guideline for good clinical practice: Note for guidance on good clinical practice, 2002.
10. Sap AG, Max. DB™—The professional DBMS; Version 1.0. Available at: <http://www.mysql.com/products/maxdb/>. Accessed: Feb 2005.
11. ICD. International classification of diseases. Available at: <http://www.who.int/classifications/icd/en/>. Accessed: Feb 2008.
12. WHO. Classification of drugs. Available at: <http://www.who.int/classifications/>. Accessed: Feb 2008.
13. The python language. Available at: <http://www.python.org/>. Accessed: Aug 2008.
14. Oracle Clinical. Available at: [http://www.oracle.com/industries/life\\_sciences/oracle-clinical.html](http://www.oracle.com/industries/life_sciences/oracle-clinical.html). Accessed: Jan 2009.
15. PhaseForward Inc. InForm™. Available at: <http://www.phaseforward.com/products/clinical/edc/>. Accessed: Jan 2009.
16. SAS. PhedIt—Solution for life science: Boosting the efficiency of clinical trial processes and product launches. Product brochure. Available at: [http://www.sas.com/offices/europe/sweden/pdf/PhedIt\\_engelska.pdf](http://www.sas.com/offices/europe/sweden/pdf/PhedIt_engelska.pdf). Accessed: Jan 2009.
17. Open Clinica. Open Source for Clinical Research. Available at: <http://www.openclinica.org>. Accessed: October 2008.
18. Nguyen L, Shah A, Harker M, et al. Dados-prospective: An open source application for web-based prospective data collection. *Sources Code Biol Med* 2006;1:7.
19. The Redcap Project. Available at: <http://www.iwg-online.org/projects/redcap/index.php>. Accessed: October 2008.